
Module 1 – Connecting to Data

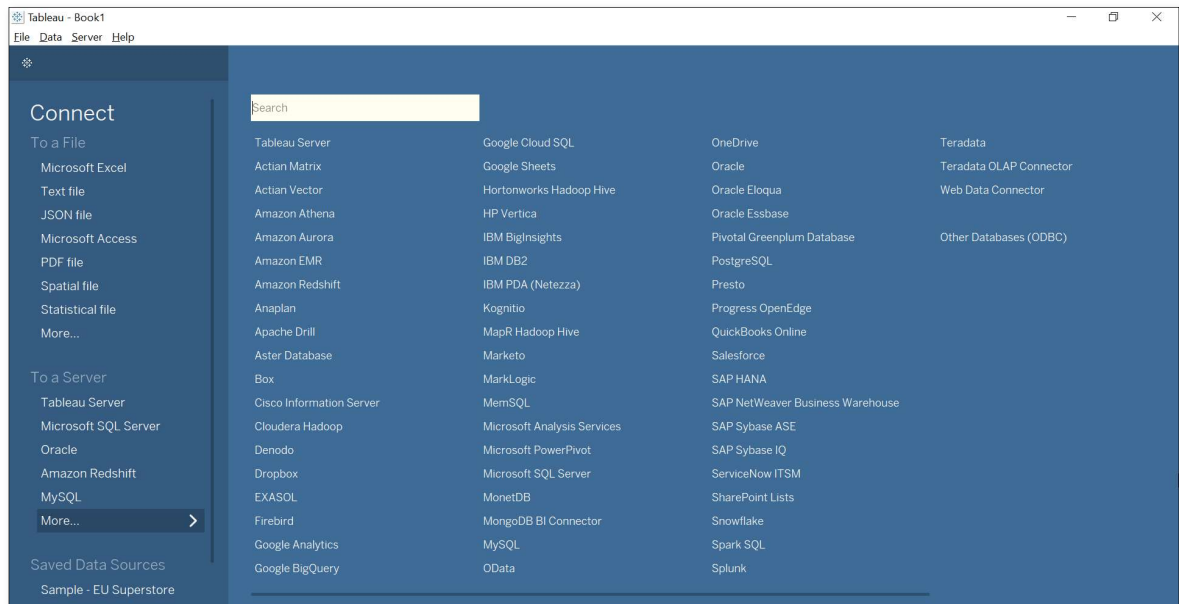
Contents:

1. Connecting to Data
2. Managing Extracts
3. Data Prep with Text and Excel Files
4. Join Types with Union
5. Cross-database Joins
6. Data Blending
7. **Saving your work!**

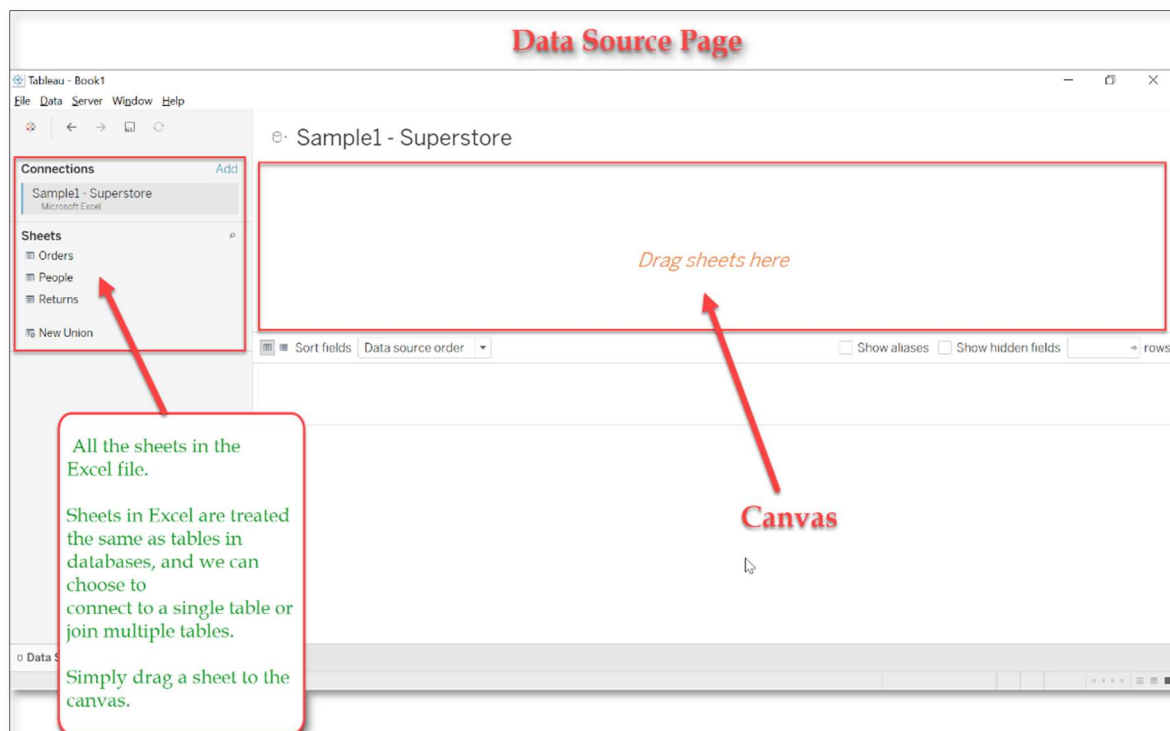
1. Connecting to Data

a. Connecting to an excel file

Tableau can connect to many data sources. In the Connect pane, we have a long list of native connections to all sorts of data sources, regardless of whether they're local files or databases on-premises or in the cloud. Example: Global Superstore Excel. Superstore is a data set of sales for a global retail chain that sells furniture, office supplies, and technology goods. Each row of data represents a single item in a transaction.



b. The data source page



Double Click to rename Data Source

Double Click to rename Table

Data Grid:

Double click column to rename

Clicking on the data type icon allows us to change the default data type for that column.

The drop down caret contains a whole menu of more advanced options.

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer ...	Segment	Country	City	State
1	CA-2013-152...	09-11-2013	12-11-2013	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentu
2	CA-2013-152...	09-11-2013	12-11-2013	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentu
3	CA-2013-138...	13-06-2013	17-06-2013	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	Califo
4	US-2012-108...	11-10-2012	18-10-2012	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderda...	Florid
5	US-2012-108...	11-10-2012	18-10-2012	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderda...	Florid
6	CA-2011-115...	09-06-2011	14-06-2011	Standard Class	BH-11710	Brosina Hoff...	Consumer	United States	Los Angeles	Califo
7	CA-2011-115...	09-06-2011	14-06-2011	Standard Class	BH-11710	Brosina Hoff...	Consumer	United States	Los Angeles	Califo

c. Changing Data Sources

Add a new Data Source

Select extract

the blue check means this is the primary connection currently being used in the viz.

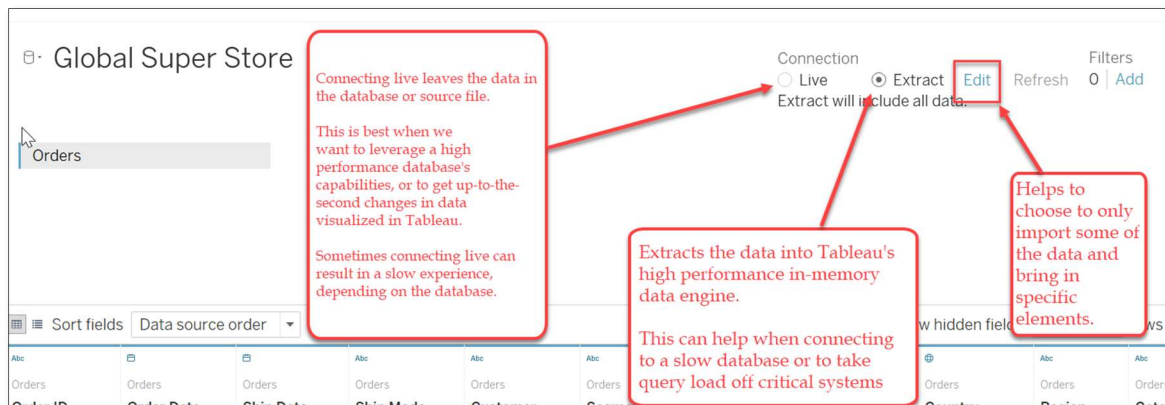
The link icons in the other data source mean we could do a data blend.

d. Switching between multiple Data Sources

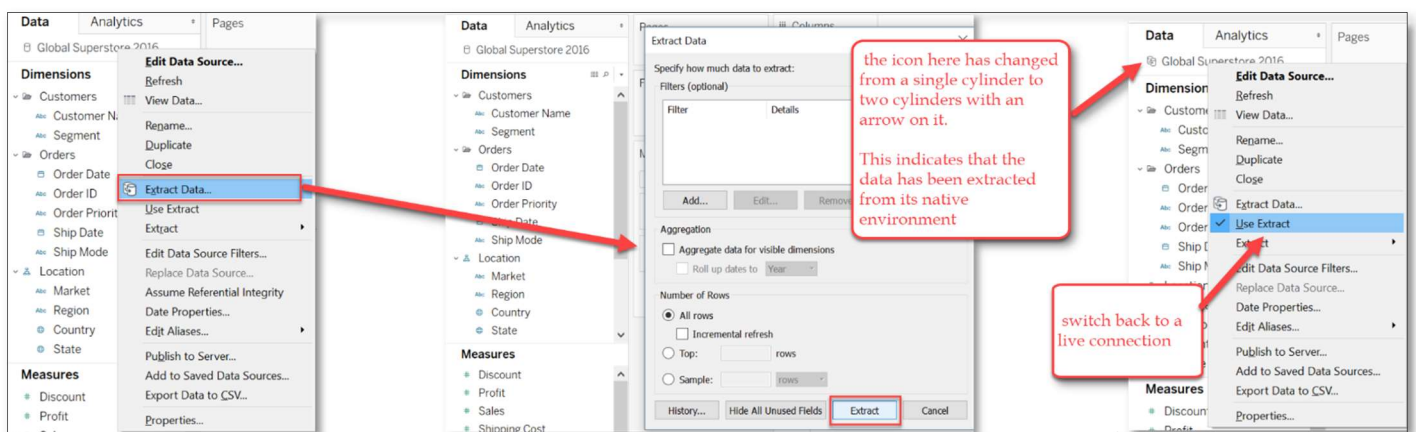
Switch between data sources

2. Managing Extracts

a. Live versus extract



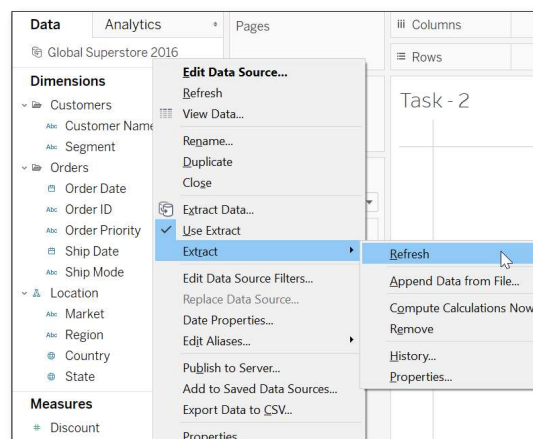
b. Creating Extracts



If the workbook has already been saved as a .twbx packaged workbook, the extract will automatically be saved as part of the packaged workbook. If the workbook has been saved as a .twb or hasn't been saved yet, we'd be asked where to save the Tableau Data Extract file (.tde) or as (.hyper).

c. Refreshing Extracts

Because we're currently connected to a static copy of the data, if the underlying data were to change, the view would not update with that new data until the extract is refreshed.



3. Data Prep with Text and Excel Files

a. Data interpreter

Now we see that those headers and nulls have been stripped out, and our columns are properly identified. We could go on, but if we want to get more specifics on what the Data Interpreter did, we can click "Review the Results". This will open an Excel file describing the changes. If we click to the tab we used, Resolved Incidents, we see which fields are being used as headers, in red, and which are considered data, in green.

Employee	01-01-2015	01-02-2015	01-03-2015	01-04-2015	01-05-2015	01-06-2015	01-07-2015	01-08-2015	01-09-2015	01-10-2015
B-002	4	1	5	2	3	0	3	1	2	0
E-055	1	2	1	3	4	1	4	0	2	0
E-075	14	17	16	15	18	16	14	17	12	0
B-066	4	4	5	2	5	0	0	2	0	0
C-025	17	13	17	18	17	17	12	15	17	0
E-030	2	2	1	1	0	3	5	5	0	0
C-001	14	14	14	14	13	18	17	14	13	0

b. Pivot

To change the format from that column-per-month layout into a single date column and a single column for Resolved Incidents, select all the date columns, open the menu and select "Pivot". This pivot feature essentially merges the information from the original columns and rows into two new columns – Pivot field names, and Pivot field values. Rename the fields to suitable names.

No. of Incidents resolved by each employee (data-wise)

Employee	01-01-2015	01-02-2015	01-03-2015	01-04-2015	01-05-2015	01-06-2015	01-07-2015
B-002	4	1	5	2	3	0	3
E-055	1	2	1	3	4	1	4
E-075	14	17	16	15	18	16	14
B-066	4	4	5	2	5	0	0
C-025	17	13	17	18	17	17	12
E-030	2	2	1	1	0	3	5
C-001	14	14	14	14	13	18	17

c. Pivot using custom SQL

d. Metadata grid

If we click on this icon here above the first column, we're brought to the metadata grid. This can be a useful view, as the vertical layout can be easier to navigate, especially if you have a large number of fields, and it's also useful when tables have been joined, or to see the original field names or even to rename the fields.

Field Name	Table	Remote Field Name
Pivot Field Names	Pivot	Pivot Field Names
Pivot Field Values	Pivot	Pivot Field Values
Employee	Resolved Incidents	Employee

e. Split

Pivot	Pivot	Resolved Incidents
Date	No of Incid...	Employee ID
01-01-2015	4	B-002
01-01-2015	1	E-055
01-01-2015	14	E-075
01-01-2015	4	B-066
01-01-2015	17	C-025
01-01-2015	2	E-030
01-01-2015	14	C-001
01-01-2015	4	E-038

Rename

Reset Name

Copy Values

Hide

Aliases...

Create Calculated Field...

Create Group...

Split

Custom Split...

Add Data to Pivot

Describe...

After Split

Pivot	Pivot	Resolved Incidents	Calculation	Calculation
Date	No of Incid...	Employee I...	Location	Employee ID
01-01-2015	4	B-002	B	2
01-01-2015	1	E-055	E	55
01-01-2015	14	E-075	E	75
01-01-2015	4	B-066	B	66
01-01-2015	17	C-025	C	25
01-01-2015	2	E-030	E	30
01-01-2015	14	C-001	C	1
01-01-2015	4	E-038	F	38

f. Custom split

Pivot	Pivot	Tiers
date	No of Incid...	Employee
01-01-2015	4	B-002
01-01-2015	1	E-055
01-01-2015	14	E-075-II
01-01-2015	4	B-066
01-01-2015	17	C-025-II
01-01-2015	2	E-030
01-01-2015	14	C-001-II
01-01-2015	4	E-038

Rename

Copy Values

Hide

Aliases...

Create Calculated Field...

Create Group...

Split

Custom Split...

Add Data to Pivot

Describe...

Options for custom split

Custom Split

How should this data be split?

Use the separator

Split off

All

1

columns

OK

Cancel

Pivot	Pivot	Tiers	Calculation	Calculation	Calculation
date	No of Incid...	Employee	Employee - ...	Employee - ...	Employee - ...
01-01-2015	4	B-002	B	002	
01-01-2015	1	E-055	E	055	
01-01-2015	14	E-075-II	E	075	II
01-01-2015	4	B-066	B	066	
01-01-2015	17	C-025-II	C	025	II
01-01-2015	2	E-030	E	030	
01-01-2015	14	C-001-II	C	001	II
01-01-2015	4	E-038	E	038	

4. Join Types with Union

Inner Join

Select all records from Table A and Table B, where the join condition is met.

Left Join

Select all records from Table A, along with records from Table B for which the join condition is met (if at all).

Right Join

Select all records from Table B, along with records from Table A for which the join condition is met (if at all).

Full Join

Select all records from Table A and Table B, regardless of whether the join condition is met or not.

Example:

Customer Table:

customer_id	first_name	last_name	email	address	city	state	zipcode
1	George	Washington	gswashington@usa.gov	3200 Mt Vernon Hwy	Mount Vernon	VA	22121
2	John	Adams	jadams@usa.gov	1250 Hancock St	Quincy	MA	2169
3	Thomas	Jefferson	tjefferson@usa.gov	931 Thomas Jefferson Pkwy	Charlottesville	VA	22902
4	James	Madison	jmadison@usa.gov	11350 Constitution Hwy	Orange	VA	22960
5	James	Monroe	jmonroe@usa.gov	2050 James Monroe Parkway	Charlottesville	VA	22902

Orders Table:

order_id	order_date	amount	customer_id
1	07/04/1776	\$234.56	1
2	03/14/1760	\$78.50	3
3	05/23/1784	\$124.00	2
4	09/03/1790	\$65.50	3
5	07/21/1795	\$25.50	10
6	11/27/1787	\$14.40	9

Note that (1) not every customer in our customers table has placed an order and (2) there are a few orders for which no customer record exists in our customers table.

Inner Join

customer

orders

Sort fields

Data source order

Show aliases

Show hidden fields

4

#	customer	customer	customer	customer	customer	customer	customer	orders	orders	orders
Project	customer	customer	customer	customer	customer	customer	customer	orders	orders	orders
Customer Id	First Name	Last Name	Email	Address	City	State	Zipcode	Order Id	Order Date	Amount
1	George	Washington	gWASHINGTON@u...	3200 Mt Vern...	Mount Vernon	VA	22121	1	07/04/1776	\$234.56
3	Thomas	Jefferson	tJEFFERSON@u...	931 Thomas J...	Charlottesville	VA	22902	2	03/14/1760	\$78.50
2	John	Adams	JADAMS@usa...	1250 Hancock...	Quincy	MA	02169	3	05/23/1784	\$124.00
3	Thomas	Jefferson	tJEFFERSON@u...	931 Thomas J...	Charlottesville	VA	22902	4	09/03/1790	\$65.50

Left Join

customer

orders

Join

Inner

Left

Right

Full Outer

Data Source

customer_id

=

customer_id

Add new join clause

Sort fields

Data source order

Show aliases

Show hidden fields

6

#	customer	customer	customer	customer	customer	customer	customer	orders	orders	orders
Project	customer	customer	customer	customer	customer	customer	customer	orders	orders	orders
Customer Id	First Name	Last Name	Email	Address	City	State	Zipcode	Order Id	Order Date	Amount
1	George	Washington	gWASHINGTON@u...	3200 Mt Vern...	Mount Vernon	VA	22121	1	07/04/1776	\$234.56
3	Thomas	Jefferson	tJEFFERSON@u...	931 Thomas J...	Charlottesville	VA	22902	2	03/14/1760	\$78.50
2	John	Adams	JADAMS@usa...	1250 Hancock...	Quincy	MA	02169	3	05/23/1784	\$124.00
3	Thomas	Jefferson	tJEFFERSON@u...	931 Thomas J...	Charlottesville	VA	22902	4	09/03/1790	\$65.50
5	James	Monroe	JMONROE@usa...	2050 James ...	Charlottesville	VA	22902	null	null	null
4	James	Madison	JMADISON@us...	11350 Consti...	Orange	VA	22960	null	null	null

Right Join

customer

orders

Join

Inner

Left

Right

Full Outer

Data Source

customer_id = customer

Add new join clause

Includes all values in right table and all matches from left table. Members without matches will show up as nulls on the left.

Sort fields

Data source order

Show aliases

Show hidden fields

6

#	Project	customer	customer	customer	customer	customer	customer	customer	orders	orders	orders
	Customer Id	First Name	Last Name	Email	Address	City	State	Zipcode	Order Id	Order Date	Amount
1	George	Washington	gWASHINGTON@...	3200 Mt Vernon...	Mount Vernon	VA	22121	1	07/04/1776	\$234.56	
3	Thomas	Jefferson	tJEFFERSON@u...	931 Thomas J...	Charlottesville	VA	22902	2	03/14/1760	\$78.50	
2	John	Adams	jADAMS@usa....	1250 Hancock...	Quincy	MA	02169	3	05/23/1784	\$124.00	
3	Thomas	Jefferson	tJEFFERSON@u...	931 Thomas J...	Charlottesville	VA	22902	4	09/03/1790	\$65.50	
10	null	null	null	null	null	null	null	null	5	07/21/1795	\$25.50
9	null	null	null	null	null	null	null	null	6	11/27/1787	\$14.40

Full Outer Join

customer

orders

Join

Inner

Left

Right

Full Outer

Data Source

orders

customer_id

=

customer_id

Add new join clause

Includes all values from both tables.
Members without matches show up as nulls.

Sort fields

Data source order

☐ Show aliases ☐ Show hidden fields 8

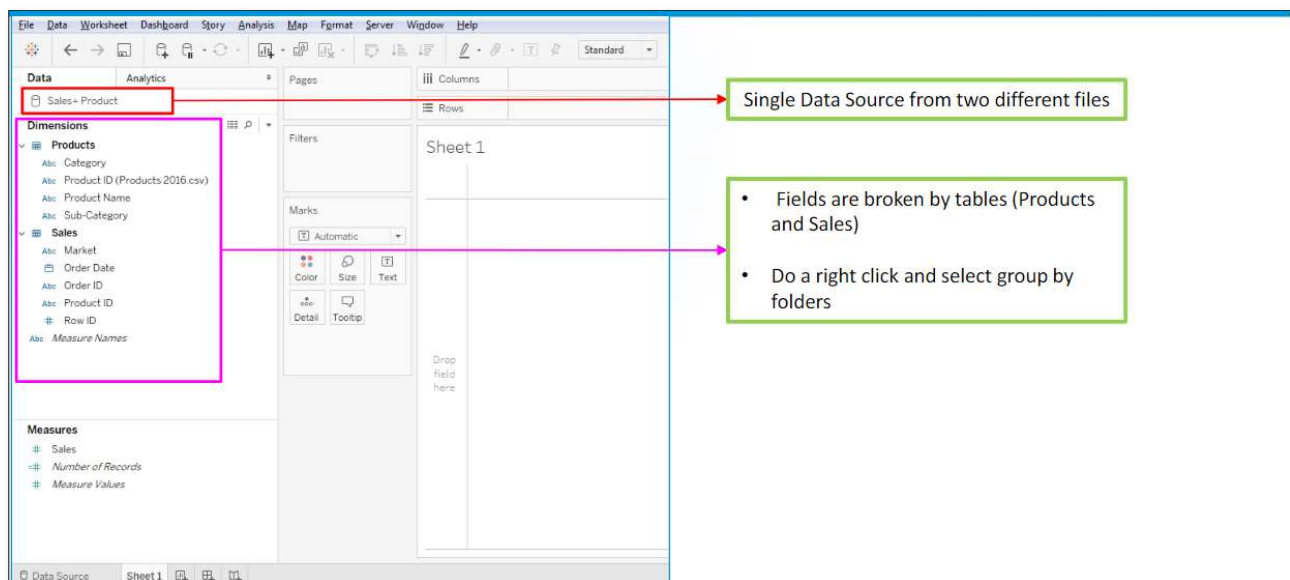
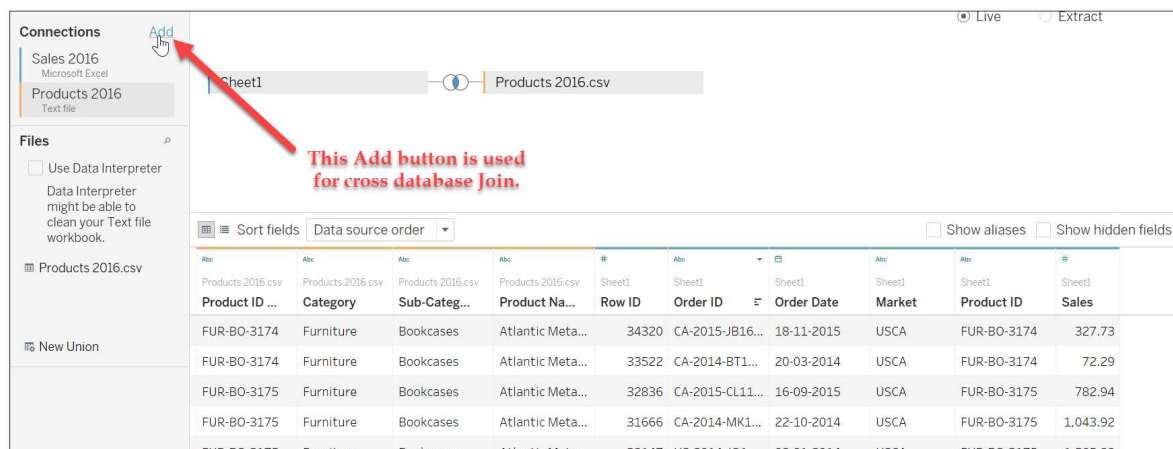
#	customer	customer	customer	customer	customer	customer	customer	#	orders	orders	orders
Customer Id	First Name	Last Name	Email	Address	City	State	Zipcode	Order Id	Order Date	Amount	
3	Thomas	Jefferson	tjefferson@u...	931 Thomas J...	Charlottesville	VA	22902	2	03/14/1760	\$78.50	
2	John	Adams	jadams@usa...	1250 Hancock...	Quincy	MA	02169	3	05/23/1784	\$124.00	
3	Thomas	Jefferson	tjefferson@u...	931 Thomas J...	Charlottesville	VA	22902	4	09/03/1790	\$65.50	
10	null	null	null	null	null	null	null	5	07/21/1795	\$25.50	
9	null	null	null	null	null	null	null	6	11/27/1787	\$14.40	
5	James	Monroe	jmonroe@usa...	2050 James ...	Charlottesville	VA	22902	null	null	null	
4	James	Madison	jmadison@us...	11350 Consti...	Orange	VA	22960	null	null	null	

Example:

Table - 1			Table - 2						
A	B	C	X	Y	A				
abc	12234	india	12	ab	xyz				
pqr	12234	india	14	cd	abc				
xyz	12234	india	15	ef	lmn				
Inner Join			Full Join						
A	B	C	X	Y	A	B	C	X	Y
abc	12234	india	14	cd	xyz	12234	india	12	ab
xyz	12234	india	12	ab	abc	12234	india	14	cd
					lmn	null	null	15	ef
					pqr	12234	india	null	null
Left Join			Right Join						
A	B	C	X	Y	A	B	C	X	Y
abc	12234	india	14	cd	xyz	12234	india	12	ab
pqr	12234	india	null	null	abc	12234	india	14	cd
xyz	12234	india	12	ab	lmn	null	null	15	ef

5. Cross-database Joins

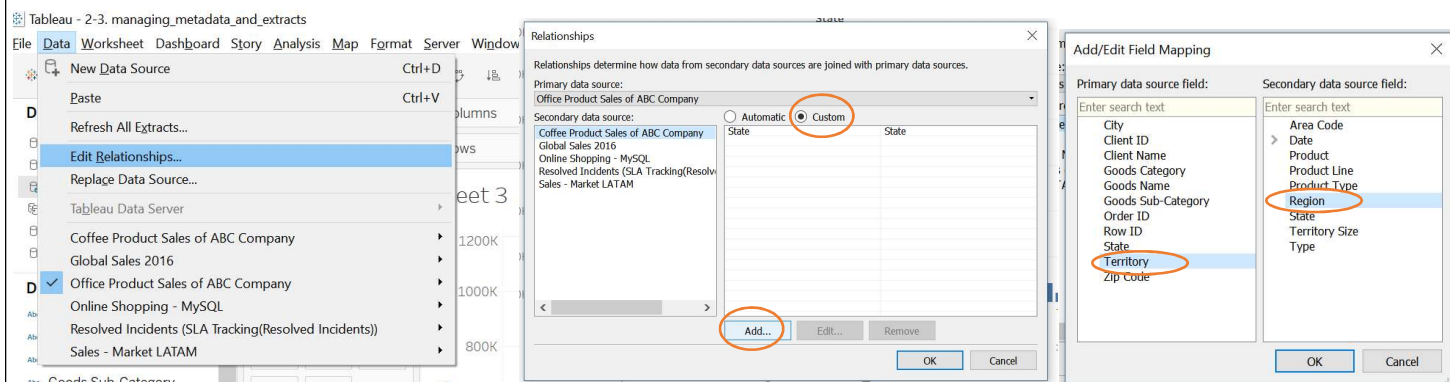
Many organizations have different data in different systems. Maybe there's financial data in a SQL Server database and product data in Amazon Redshift. The data may live in different environments, but we want to do analysis on everything together.



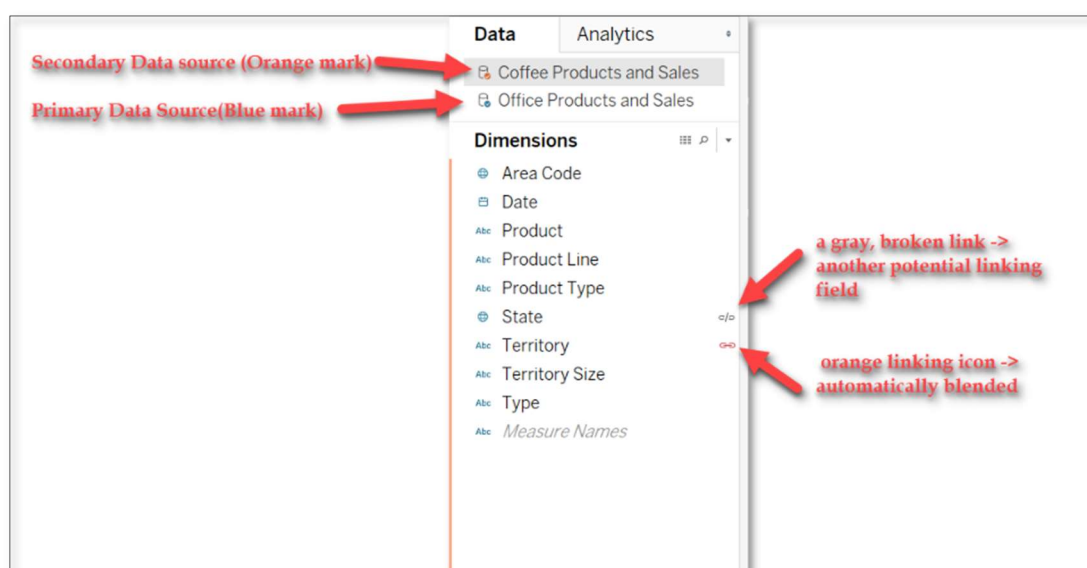
6. Data Blending

Imagine you own two retail chains, Office Products and Coffee Products. You capture your Office Products data in one system and your Coffee Products data in another. Data blending is one way of combining data from multiple data sources into a single view. Instead of joining the data at the row level like a cross-database join, data blending sends separate queries to the separate data sources and aggregates the results to a common level back in Tableau. Data blending requires at least one common field between both data sources.

Example: *State* is common field in our example. *Region* from Office products and *Territory* from Coffee Products refer to same data. We can match it together either by renaming one of the field or by “Data > Edit Relationships” and the adding a custom relationship between them. If they have the same name Tableau can create a relationship between both data sources automatically.



Establishing relationship between Region from Office products and Territory from Coffee Products.



It's important to note that primary and secondary sources are determined on a worksheet-by-worksheet basis and are not maintained globally throughout the workbook. When we are on a new sheet, the data sources within the Data Pane do not have orange and blue check marks to indicate them as primary and secondary. The relationships we established in the previous worksheet are not carried over.

What happens when we swap our data sources and use Coffee Products as the primary source and Office Products as the secondary? We'll drag Coffee Products Sales to the view first, then we'll bring out State. There are now only about half the states on the view compared with our previous example. This is because Coffee Products has fewer states in its data set than Office Products. Tableau is displaying all the states in Coffee Products, and then it will pull in only the relevant information from Office Products. This means that any state in Office Products that is not in Coffee Products will not be displayed in this view.

7. Saving your work!

- **Workbooks (.twb)** – Tableau workbook files have the .twb file extension. Workbooks hold one or more worksheets, plus zero or more dashboards and stories.
- **Bookmarks (.tbn)** – Tableau bookmark files have the .tbn file extension. Bookmarks contain a single worksheet and are an easy way to quickly share your work.
- **Packaged Workbooks (.twbx)** – Tableau packaged workbooks have the .twbx file extension. A packaged workbook is a single zip file that contains a workbook along with any supporting local file data and background images. This format is the best way to package your work for sharing with others who don't have access to the original data.
- **Extract (.hyper or .tde)** – Depending on the version the extract was created in, Tableau extract files can have either the .hyper or .tde file extension. Extract files are a local copy of a subset or entire data set that you can use to share data with others, when you need to work offline, and improve performance.
- **Data Source (.tds)** – Tableau data source files have the .tds file extension. Data source files are shortcuts for quickly connecting to the original data that you use often. Data source files do not contain the actual data but rather the information necessary to connect to the actual data as well as any modifications you've made on top of the actual data such as changing default properties, creating calculated fields, adding groups, and so on.
- **Packaged Data Source (.tdsx)** – Tableau packaged data source files have the .tdsx file extension. A packaged data source is a zip file that contains the data source file (.tds) described above as well as any local file data such as extract files (.hyper or .tde), text files, Excel files, Access files, and local cube files. Use this format to create a single file that you can then share with others who may not have access to the original data stored locally on your computer.